Bellabeat Business Analysis

Ask phase:

Problem statement:

Challenge: Understand how consumers utilize non-Bellabeat smart devices to identify opportunities to improve Bellabeat's marketing strategy and potentially the functionalities of a specific Bellabeat product.

Analysis Goals:

1. Identify key trends in user behavior related to non-Bellabeat smart devices. This may involve aspects like frequency of use, preferred functionalities, device categories (fitness trackers, sleep monitors, etc.), and potential pain points or gaps users experience with existing devices.
2. Analyze how these trends can be applied to understand Bellabeat's customer base. This could involve exploring potential overlaps between Bellabeat users and users of specific non-Bellabeat devices, identifying unmet needs of Bellabeat users based on the broader smart device usage trends, and understanding how Bellabeat can position itself within the competitive landscape.
3. Develop high-level marketing recommendations for Bellabeat based on the insights from the analysis. This might suggest targeting specific customer segments, emphasizing functionalities that address identified user needs, or exploring potential product improvements based on broader smart device trends.

Business task

Objective: Analyze user data from a specific Bellabeat smart device to understand user behavior and identify potential areas for improvement in the product and marketing strategy.
Data Source: User data from the chosen Bellabeat smart device, including metrics like activity levels, sleep patterns, stress levels, and any other relevant data points.
Deliverables:

- **Analysis:** Identify key trends and insights into how users are interacting with the device. This may include user demographics, typical usage patterns, areas where users might be disengaged, and potential correlations between different data points (e.g., sleep and activity levels).

- **Visualization:** Create clear and concise visualizations (charts, graphs) to effectively communicate findings to the executive team.
- **Marketing Recommendations:** Based on analysis, propose high-level marketing strategies to improve user engagement, address identified needs, and potentially attract new customer segments.

Target Audience: Bellabeat executive team

Prepare phase:

Data source

FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius): This Kaggle data set contains personal fitness trackers from thirty Fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users' habits.

Data storage

The factors for storage are:
- **Security**: the dataset has sensitive information about users like daily activities, sleeping schedule, stress, and reproductive health it is of great concern to respect the user's privacy and keep the data as secure as possible.
- **Accessibility**: a marketing analytics team is working on this dataset to gain some powerful insight from the dataset and make data-driven business so it needs to be accessed by the team.
- **Collaboration**: there are multiple members working on the same dataset and that should be reflected to other members of the team in real-time so the platform for storage should be collaborative.

Considering the following factors for storage **Cloud-based storage** is best suited to fulfill the purpose.

Data Organization

A wide data format should be considered for analysis because of the following reasons:

- **Intuitive Analysis:** Each row represents a single user, and each column represents a specific data point (e.g., steps on Day 1, heart rate on Day 2). This makes it easier to understand user behavior across different metrics and compare data points for individual users.

- **Grouping and Aggregation:** A wide format allows users to be easily grouped by demographics or usage patterns and calculate aggregate statistics (e.g., average daily steps, sleep duration) for each group. This facilitates identifying trends within user segments.
- **Time Series Analysis:** the data includes timestamps for each data point (e.g., hourly step count), a wide format is ideal for time series analysis.

Bias or credibility of data

There can be the following types of biasing in the dataset:

- **Sample Size:** The dataset only includes 30 Fitbit users. This is a relatively small sample size and might not be representative of the entire Fitbit user base or the broader population of smart device users. Be mindful of generalizing your findings from this limited data set.
- **Selection Bias:** The data comes from users who consented to share their data. These users might be more health-conscious or engaged with fitness tracking compared to the general population.
- **Device Specificity:** The data is specific to Fitbit devices. User behavior with other smart device brands might differ.

Data's Integrity

Accessibility:

- **Open access:** The data is freely available and unrestricted by copyright, to use it for analysis without permission hurdles.

Reliability :

- **Data source transparency:** While the data is accessible, its origin remains unclear. It was collected through a third-party platform (Amazon Mechanical Turk) which can introduce uncertainty about the participants and data collection methods.

Comprehension:

- **Limited metadata:** The original dataset lacks information about data collection methodology and measurement units. This can make interpretation challenging.
- **Complementary resource:** Fortunately, another Kaggle user provided a separate metadata file specifically for the Fitbit data. This metadata can serve as a valuable guide to understanding the data collection process and interpreting the values.

Potential for Selection Bias:

- The data comes from users who consented to share their fitness tracking information. These users might be more health-conscious or engaged with fitness tracking compared to the average person. This selection bias could skew the results towards more active users. Consider this potential bias when interpreting the findings.

Data Specificity to Fitbit Devices:

- The data is specific to Fitbit devices. User behavior with other smart device brands (different functionalities, interfaces, etc.) might differ significantly. Be cautious about extrapolating findings from this dataset to users of non-Fitbit devices.

Can this data help in answering questions?

With a bit of data cleaning, data manipulation, and organization this data can help in answering the questions in business tasks.

Process phase:

Tool for analyzation

Factors for choosing tools for analysis:
- Handle large dataset
- Collaborative
- Support both visualization and manipulation of data

Considering the above factors the R language has all the features to meet the requirements for current analysis.

Ensuring data integrity

Data integrity has been ensured by initial inspection in R using str() and summary() functions, checking the correlation between the fields of the dataset, and identifying and fixing the missing values, null values, and redundant values.

Steps that were taken to ensure clean data

1. Initial inspection using str() and summarize() to check the data type of field and their values.

2. Fixing missing values
3. Correcting the format for entered data
4. Removing redundant values
5. Arranging the values according to need.

Documentation of data cleaning

Documentation of data cleaning has been created so that it can be reviewed and shared among other team members.

Analyze phase:

Summary of different parameters:

calorie summary

```
##     Calories
## Min.   :   0
## 1st Qu.:1828
## Median :2134
## Mean   :2304
## 3rd Qu.:2793
## Max.   :4900
```

intensity summary

```
## VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
## Min.   :  0.00    Min.   :  0.00      Min.   :  0.0        Min.   :   0.0
## 1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:127.0        1st Qu.: 729.8
## Median :  4.00    Median :  6.00      Median :199.0        Median :1057.5
## Mean   : 21.16    Mean   : 13.56      Mean   :192.8        Mean   : 991.2
## 3rd Qu.: 32.00    3rd Qu.: 19.00      3rd Qu.:264.0        3rd Qu.:1229.5
## Max.   :210.00    Max.   :143.00      Max.   :518.0        Max.   :1440.0
```

total step summary

```
##    StepTotal
## Min.   :    0
## 1st Qu.: 3790
## Median : 7406
## Mean   : 7638
## 3rd Qu.:10727
## Max.   :36019
```

```
##  TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
 ## Min.   :1.000    Min.   : 58.0    Min.   : 61.0
 ## 1st Qu.:1.000    1st Qu.:361.0    1st Qu.:403.0
 ## Median :1.000    Median :433.0    Median :463.0
 ## Mean   :1.119    Mean   :419.5    Mean   :458.6
 ## 3rd Qu.:1.000    3rd Qu.:490.0    3rd Qu.:526.0
 ## Max.   :3.000    Max.   :796.0    Max.   :961.0
```

Findings from statistics summary

- Calorie The number of calories you burn each day depends on your individual goals. To maintain your weight, you need to burn roughly the same amount of calories you consume. Without considering your specific goals, it's impossible to determine if you're burning enough calories.
- Intensity The Centers for Disease Control and Prevention (CDC) recommends at least 30 minutes of moderate-intensity exercise daily. The data is positive, showing the participants averaged 34.72 minutes of activity (combining very active and fairly active periods). This suggests they are meeting or exceeding the CDC's recommendation. The average sedentary participants were 991.2 minutes or 16.52 hours. More than half of the participants spend their time with no physical activity!
- Step While the participants average 7638 daily steps, falling short of the CDC's recommended 8,000 steps for adults, a 5% daily increase would bridge the gap.
- Sleep The average participant sleeps for 7 hours a day.

Total step vs total calories graph

Total Steps vs. Calories

Calorie

The analysis reveals a moderate correlation (coefficient of 0.59) between step count and calories burned. While Fitbit displays a calorie/step ratio, this offers a limited view because other factors can significantly impact calorie expenditure.

Graph of different categories of participants

Participants Distribution by Active Minutes
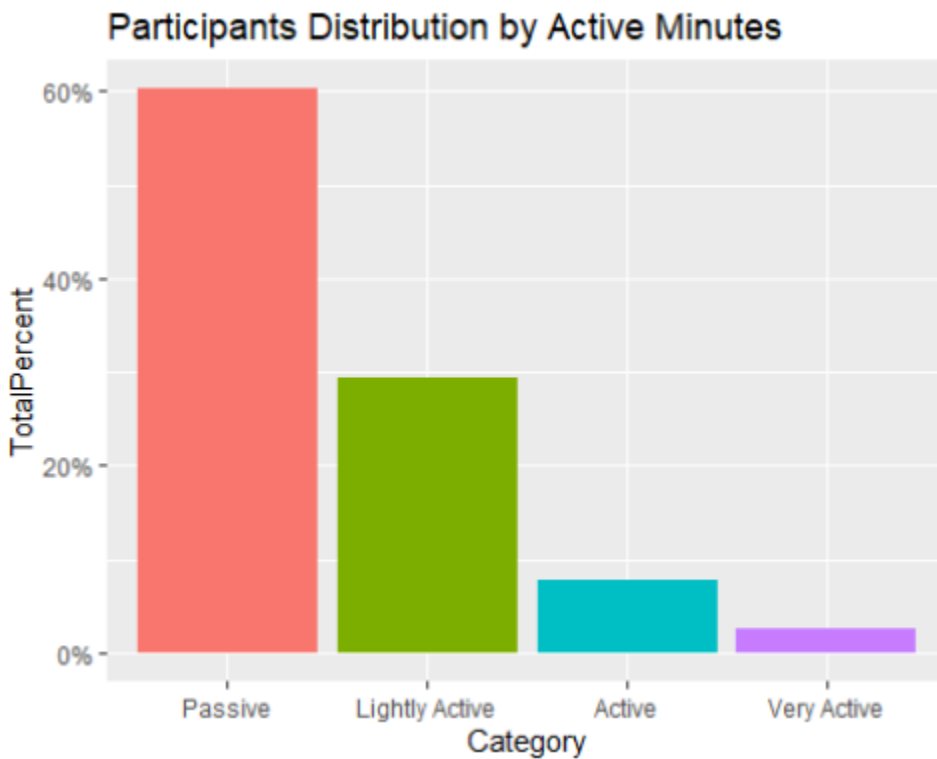
Intensity

The column graph reveals significant discrepancies between participant activity levels. Notably, around 60% fall within the "Passive" category. To delve deeper into this, we'd need more granular data on participants' activities, ideally broken down by hours spent in each category.

Graph of average total steps vs time

Average Total Steps vs. Time
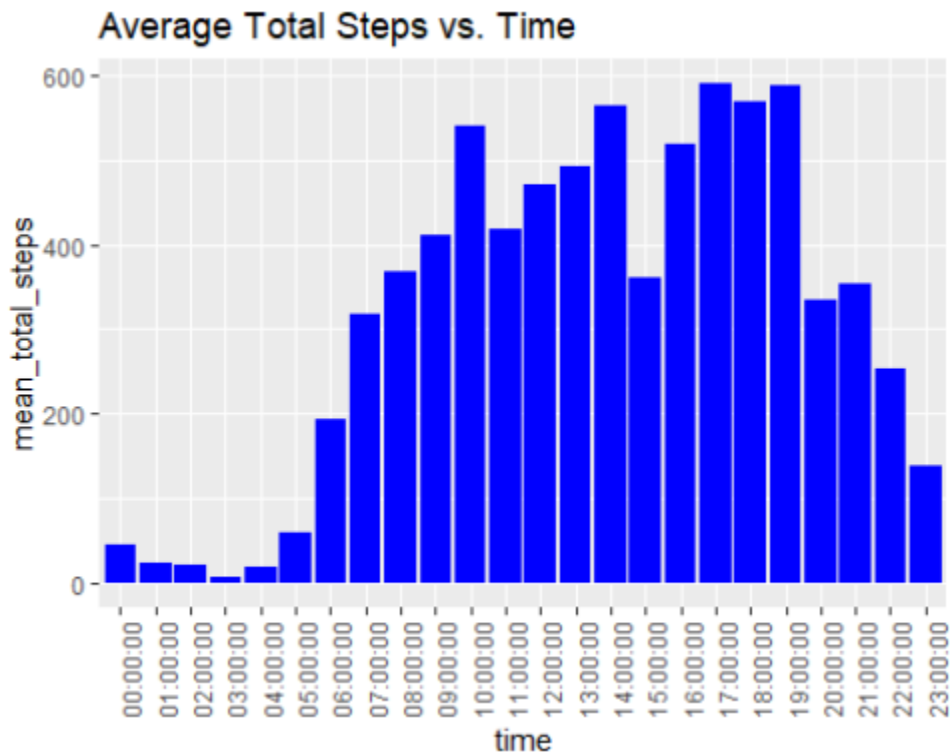
The histogram reveals interesting patterns in participant activity throughout the day. Steps are highest in the morning around 9 AM, potentially due to commuting to work on foot. Activity picks up again during lunchtime (12 PM - 1 PM), suggesting movement during work breaks. Finally, the daily peak occurs between 5 PM and 7 PM, coinciding with the end of the workday. This evening surge likely reflects exercise routines like walking, running, or gym sessions.
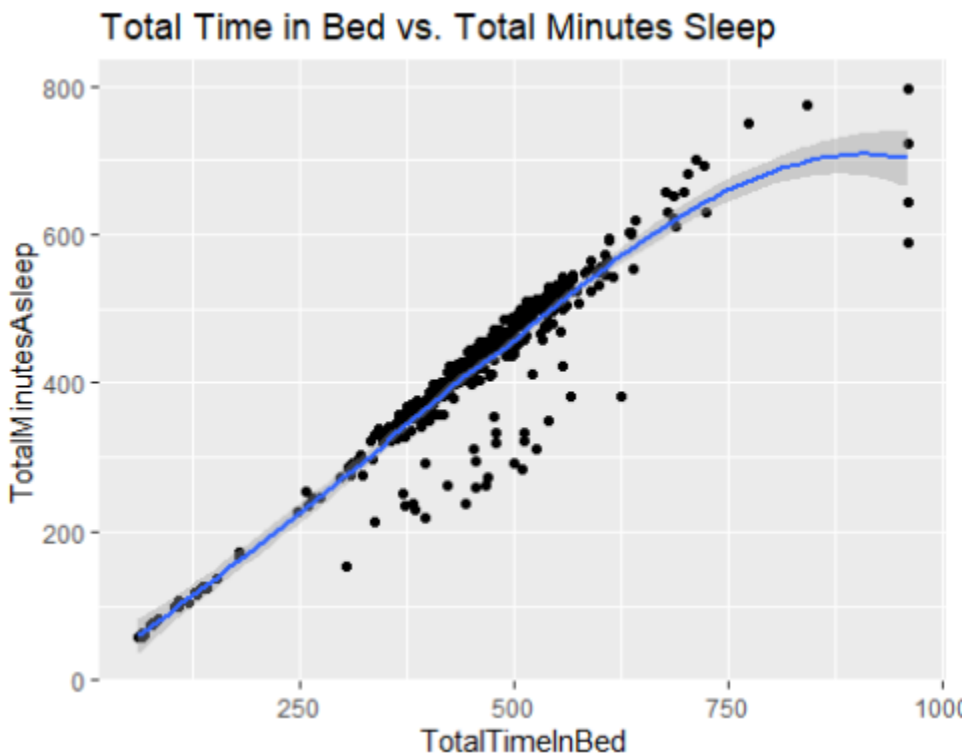
Graph for participants' distribution by sleep time

Participants Distribution by Sleep Time
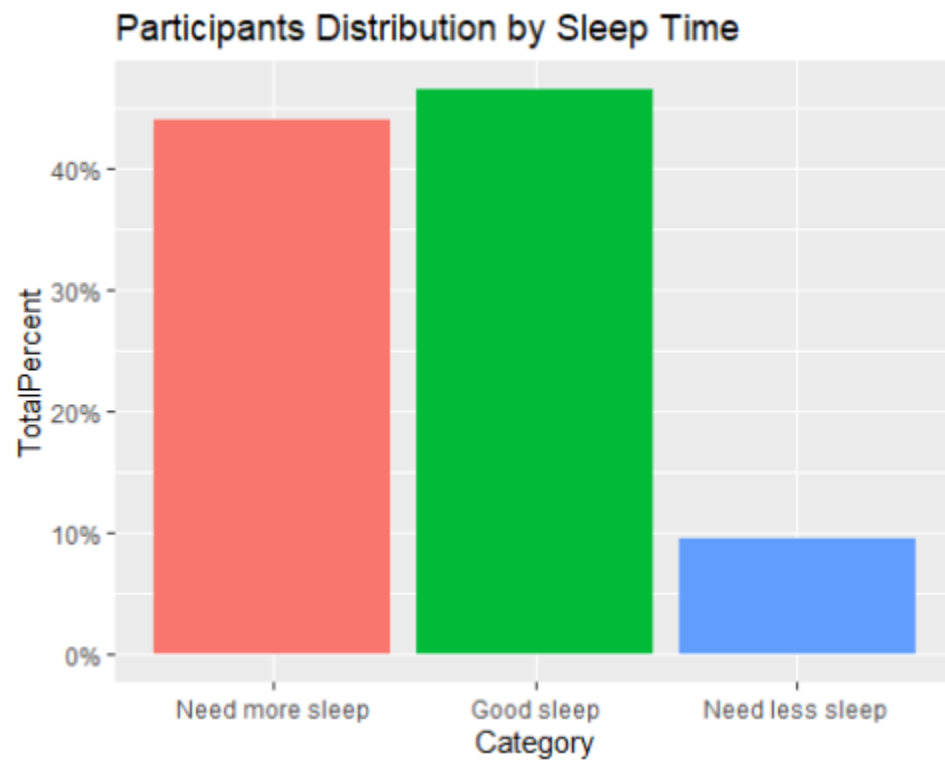
Sleep Time vs. Time in Bed Correlation

[1] 0.9304575

Graph of total time in bed vs total minutes in sleep

Sleep

While both the CDC (at least 7 hours) and Sleepfoundation.org (less than 9 hours) recommend optimal sleep ranges for adults, only 46% of participants appear to be achieving this based on the data. To improve sleep habits, implementing a sleep reminder notification and built-in alarm features could be valuable additions to engage users.

The analysis reveals a strong correlation (r = 0.93) between total time in bed and total sleep minutes. However, some participants show a significant difference between these two measures. Highlighting this data for users could be beneficial. For example, users with a pattern of large discrepancies between time in bed and sleep duration could be prompted to utilize the "Reminder to Go to Sleep" feature if they're aiming to increase their total sleep time.

Recommendation¶

Fitbit's data offers valuable insights into user activity levels, but some limitations exist. To get a complete picture of calorie expenditure, users need to track their daily calorie

intake as well. Fitbit could integrate a food/drink logging feature with average calorie information to address this.

The data also reveals an average of 16.5 hours of sedentary time per day. To encourage movement, Fitbit could implement gentle reminder nudges throughout the day. However, these reminders shouldn't disrupt work. To achieve this, Fitbit could introduce a customizable "Work Time" setting where reminders are paused

While average daily steps fall short of the CDC's recommended 8,000, Fitbit's "Goal Level" feature can help users bridge the gap. This feature provides tiered daily step goals, like Passive (2,500 steps), Lightly Active (5,000 steps), and Active (at least 8,000 steps). When users consistently surpass a level, Fitbit can suggest a "level up" to the next challenge, nudging them to gradually increase their daily steps and become more active.

The data highlights that over half of participants are falling short of recommended sleep durations. Additionally, some individuals show a considerable discrepancy between time spent in bed and actual sleep time. To address these findings, Fitbit could:

-Display daily sleep totals upon waking: This would provide users with immediate feedback on their sleep performance.

-Utilize sleep reminders: Users aiming to increase sleep duration could leverage customizable bedtime reminders.

-Leverage smart alarms: For those seeking to achieve a specific sleep target, integrating alarms that trigger upon reaching that desired sleep duration could be beneficial.